

## O QUE PODEMOS ESPERAR DA ANÁLISE DE DADOS QUALITATIVOS SUPORTADA POR SOFTWARE?

António Pedro Costa

António Pedro Costa<sup>1</sup> e Maria Cecília de Souza Minayo<sup>2</sup>

<sup>1</sup>Ludomedia/webQDA e Centro de Investigação de Didática e Tecnologia na Formação de Formadores Research (CIDTFF), Universidade de Aveiro, Portugal e <sup>2</sup>Fundação Oswaldo Cruz, Brasil

apcosta@ua.pt; maminayo@terra.com.br

### Resumo

A análise de dados qualitativos assistida por computador (CAQDAS) tem aumentado em termos de procura, mas também nas soluções disponíveis para os investigadores. O cenário atual coloca essas ferramentas como essenciais, não só porque fornecem precisão e sistematização para projetos de pesquisa, mas também porque permitem alcançar resultados que, de outra forma, não seria possível. Neste contexto, os pacotes de *software* incorporam funcionalidades que lhes permitem “adaptar-se” a diferentes técnicas de análise. Neste ensaio, procura-se refletir sobre os caminhos que estas ferramentas estão a ou poderão “pisar”. De forma análoga ao uso específico de ferramentas para a análise de dados, nenhum leitor acreditaria que este texto foi escrito com uso de uma máquina de datilografar.

**Palavras-chave:** CAQDAS. Investigação Qualitativa. Investigação Colaborativa. Processamento e Transcrição Automática.

### Abstract

Computer Assisted Qualitative Data Analysis (CAQDAS) has increased in terms of demand, but also in the solutions available to researchers. The current scenario places these tools as essential, not only because they provide accuracy and systematization for research projects, but also because they allow for results that otherwise would not be possible. In this context, software packages incorporate features that allow them to “adapt” to different analytical techniques. In this essay, we try to reflect on the paths that these tools are or can following. Analogously to the specific use of tools for data analysis, no reader would believe that this text was written using a typewriter.

**Keywords:** CAQDAS. Qualitative Research. Collaborative Research. Automatic Transcription and Processing.

### ANÁLISE DE DADOS SUPORTADA POR SOFTWARE

Este texto apresenta uma reflexão baseada em publicações do próprio autor e nos estudos que tem realizado enquanto coautor/investigador do software de análise dados qualitativos webQDA (SOUZA; COSTA; MOREIRA, 2016).

Os pacotes de *software* específicos de suporte à investigação qualitativa possibilitam a organização e sistematização da recolha e análise dos dados, bem como potenciam a definição de dimensões, categorias e subcategorias de análise, processos usualmente muito laboriosos (SOUZA; COSTA; SOUZA, 2015). Por outro lado, a investigação qualitativa produz

frequentemente uma quantidade elevada de dados que exigem “organização, estruturação e redução sem prejuízo da qualidade das inferências que se procura produzir. O rigor deve pautar o momento de tratamento e interpretação de dados, devendo o investigador qualitativo socorrer-se de todas ferramentas disponíveis para assegurar a qualidade do seu trabalho, como o uso de *software* dedicado, tal como fazem aqueles que recorrem à estatística inferencial para a comprovação de hipóteses.” (RIBEIRO; BRANDÃO; COSTA, 2016, p. 158)

Os diferentes pacotes de *software* têm-se munido de novas funcionalidades com a pretensão de dar resposta às diversas metodologias e técnicas de análise de dados. Por outro lado, autores como Kaefer, Roper & Sinha (2015) referem que muitos investigadores discutem as vantagens e limitações do uso de pacotes de *software* na análise qualitativa de dados na investigação. No artigo “*A Software-Assisted Qualitative Content Analysis of News Articles: Example and Reflections*” estes autores afirmam que o uso de *software* torna mais rápida e eficiente a gestão de dados, a gestão de ideias, permite modelar graficamente essas ideias e analisar um maior volume de dados. Ao mesmo tempo, reduz a complexidade do processo de análise, potencia o rigor metodológico, a consistência e transparência analítica. Além disso, as funcionalidades destes pacotes de *software* possibilitam a procura e a modelação dos dados, dando-lhes uma visibilidade que não seria possível com métodos manuais, permitindo, assim, obter novas percepções (*insights*) e reflexões sobre um determinado projeto/estudo. Por outro lado, alguns pacotes de *software* exigem um (elevado) investimento financeiro e de tempo na sua aprendizagem.

Muitos utilizadores confiam em demasia nestes pacotes que, muitas vezes, criam expectativas irrealistas. Bazeley (2007) refere que a relativa facilidade de codificação assistida por *software* pode reduzir a leitura crítica e reflexiva e que irá mecanizar a análise qualitativa e assim comprometer o carácter exploratório e interpretativo da maioria das investigações qualitativas. Richards (2002) afirma que o excessivo uso de *software* pode levar ao “fetichismo de codificação” [nossa tradução de “coding fetishism”]. Porém, tal como atualmente nenhum investigador equaciona escrever um artigo usando uma máquina de dactilografar, também neste domínio há que saber aliar a sabedoria e a prudência do artesão que questiona, fundamenta, decide e age com pessoal e experiencial domínio da matéria que trabalha, com as potencialidades das novas tecnologias.

No que concerne à análise de dados qualitativos, as atuais ferramentas permitem analisar diferentes tipos de dados (texto, imagem, áudio e vídeo), explorar diversos métodos e técnicas de análise, codificar e categorizar, interpretar e inferir os dados. Cabe ao investigador decidir como e quais os dados a codificar e garantir que o uso do *software* se adapte ao quadro teórico, ao contexto e às questões de investigação.

Este texto está dividido em três partes: a primeira, agora finda, que faz uma pequena abordagem aos designados CAQDAS; a segunda, sobre os possíveis caminhos que estas ferramentas estão ou poderão optar; a terceira e última parte com algumas considerações sobre as vantagens e desvantagens do uso de ferramentas específicas para a análise de dados.

## 2. TRÊS POSSÍVEIS CAMINHOS

Os investigadores e decisores deparam-se, atualmente, com o desafio de gerir grandes quantidades de informação, em formatos cada vez mais diversos. Os pacotes de *software* de análise de dados têm a função de facilitar, otimizar e agilizar o tratamento desta informação, tornar os resultados mais rápidos e eficazes. A implementação de funcionalidades avançadas, integrando os conceitos de *Cloud*, *Big Data* e tecnologias móveis para promover a análise de dados na investigação qualitativa em redes internacionais de investigação é uma realidade, mas também em investigações locais e individuais.

Os chamados *Computer Assisted Qualitative Data Analysis* (CAQDAS) têm evoluído nas últimas décadas, em várias dimensões. Consideramos que os dois principais marcos desta evolução foram: i) possibilidade de trabalhar diversos formatos de dados, partindo do “somente texto” para os atuais dados multimédia (imagem, áudio e vídeo) e outros formatos, tais como: .pdf, .xml, .xls etc., e ii) integração de processos de visualização e das mais-valias da internet.

No entanto, existem muitos outros desafios que os CAQDAS necessitam de enfrentar e que podem ser resumidos a três principais áreas (COSTA; AMADO, 2018):

- i) **transcrição automática e integrada de dados multimédia:** a transcrição automática é um anseio dos investigadores, profissionais e cientistas sociais nesta e noutras áreas do conhecimento. Além de transcrições automáticas e flexíveis, os pacotes de software devem permitir aos investigadores validar, de forma integrada, os demais processos de análise de dados e os resultados requeridos. Falamos, por um

lado, de uma necessidade de automatização e, por outro, da importância de tornar mais rigorosos os processos de validação e de análise com base na inteligência humana. Atualmente, os profissionais que trabalham com a investigação de dados multimídia despendem muito tempo para obter uma primeira versão transcrita dos dados onde possam ler e refletir antes da sua codificação e síntese final. Mesmo utilizando processos de transcrição humana, de custo elevado, manda o rigor metodológico que estas transcrições sejam validadas pelos autores originais dos discursos e, na sua impossibilidade, por outros investigadores igualmente idôneos. Assim, os processos de auditoria e validação são inerentes às metodologias qualitativas e devem ser executados independentes da origem das primeiras versões das transcrições, sejam estas humanas ou de máquina. A partilha do texto transcrito, devidamente indexado a partes específicas e correspondentes dos trechos de vídeo ou do áudio, permitirá a sua correção em tempo real por todo um grupo de investigadores que esteja a trabalhar de modo colaborativo num projeto. Estas soluções necessitam de estar articuladas com outras partes de um CAQDAS, permitindo a busca da coerência interna do projeto, desde a transcrição automática dos dados, à sua validação, codificação e síntese final dos resultados.

- ii) **processamento automático e integrado de códigos:** existiriam vantagens em se ter uma Web Semântica, ou seja, um processo de automação ou inteligência artificial aplicadas à codificação de dados nos pacotes de *software* de análise qualitativa. Consideramos que os computadores não colocam em causa o trabalho dos cientistas sociais, mas antes o potenciam. Igualmente, acreditamos que a incorporação de processos de automação nesta área específica não colocará em causa o trabalho dos seus profissionais, assim como a automação na leitura de imagens radiológicas não coloca em causa o trabalho dos médicos. Pelo contrário, este tipo de tecnologia impulsiona estes especialistas para outros níveis de competências e reflexão, ao proporcionar rapidez, rigor e produtividade. Analisando, por exemplo, e especificamente, o processo de codificação de dados não-numéricos e não-estruturados numa lógica categorial, é pressuposto metodológico de rigor e coerência científica que o sistema de análise e sua codificação seja validado por outros investigadores, mesmo que a primeira versão da codificação tenha sido feita pelo ser

humano. No caso de esta primeira etapa de codificação ser feita por algum processo de automação ou inteligência artificial, será sempre necessário um processo de auditoria ou validação por investigadores. Além deste processo de validação, todas as etapas de inferência, síntese, determinação das implicações e contextualizações dependerão do fator humano. Assim, compreendemos que os processos de automação e inteligência artificial, articulados com técnicas de Web Semântica, *Big Data* etc. poderão potencializar os processos de codificação dos CAQDAS, especificamente no uso das técnicas de análise de dados.

- iii) **otimização dos processos de colaboração *online*:** disponibilizar uma ferramenta de análise qualitativa para os investigadores e não pensar em funcionalidades que permitam a colaboração, seria contrariar a tendência atual. O trabalho de investigação não é solitário e a sua qualidade pode ser favorecida com esta partilha. Desde a segunda metade do século passado, os suportes tecnológicos e mediáticos têm contribuído tanto para a organização e análise de dados, como para a produção colaborativa do conhecimento nos diversos contextos investigativos. Os CAQDAS devem ampliar as possibilidades de comunicação entre os investigadores na definição das dimensões de análise (categorização, codificação e recodificação, etc.), fundamentais para as construções de síntese e de análise (COSTA; LINHARES; SOUZA, 2014). Os pacotes de *software* colaborativos surgiram essencialmente para apoiar o desenvolvimento de aplicações informáticas. As atuais exigências no contexto investigativo compelem cada vez mais os investigadores a apetrecharem-se de ferramentas digitais que proporcionem a agilização e eficiência nos seus processos de investigação. Independentemente da natureza da pesquisa, o investigador ao recorrer às ferramentas digitais procura assegurar que os dados por si recolhidos sejam analisados de forma criteriosa e sistemática, gerindo mais eficazmente o tempo despendido e aumentando a fiabilidade dos resultados obtidos (BAUGH; HALLCOM; HARRIS, 2010). Por outro lado, a integração de *software*, especialmente o baseado em *cloud computing*, veio agilizar o trabalho colaborativo entre investigadores no processamento de dados qualitativos, um trabalho tradicionalmente solitário, assente em posterior validação por pares (COSTA; SOUZA; SOUZA, 2016). Os pacotes de *software* colaborativo surgiram



essencialmente para apoiar o desenvolvimento de aplicações informáticas. Por exemplo, estes pacotes permitem planificar, definir o tempo estimado, indicar o tempo gasto, atribuir tarefas pelos diferentes membros, construir um plano de Gantt, oferecer uma biblioteca para partilha de documentos, wiki, FAQ e ferramentas gerais de gestão (acompanhamento, monitorização, gestão de alterações, gestão de reuniões, entre outras). Esta realidade não tem sido amplamente explorada pelos CAQDAS, apesar de facilitar partilha de informação e a comunicação entre os elementos de uma equipa (SERÇE et al., 2010). É possível, assim, num espaço online, disponibilizar um conjunto de ferramentas, que podem ser agrupadas, de acordo com as suas funcionalidades genéricas, nas seguintes categorias: sistemas de comunicação (fóruns, chats); espaços de partilha de informação (*mediaspaces*); e coordenação de processos (*workflow*).

Embora estes três domínios ainda sejam vistos com alguma expectativa no contexto da investigação qualitativa e tragam desafios tecnológicos consideráveis, compreendemos que atualmente existem condições para avançar nestas inovações. A massa crítica atual, com uma comunidade académica cada vez mais “tecnológica” e com tecnologias disponíveis cada vez mais “maduras”, admite que sejamos otimistas na implementação de novas funcionalidades que permitam fazer face aos desafios apresentados. Destacamos de seguida algumas problemáticas relativas a cada uma das referidas áreas.

### 3. ALGUMAS CONSIDERAÇÕES

Grande parte da comunidade científica acredita que são inúmeras as vantagens no uso de ferramentas específicas na análise de dados: a) permitem a análise de uma enorme quantidade de dados; b) incluem procedimentos de contextualização e de validação; c) permitem a definição de categorias de forma indutiva e dedutiva, codificação e recodificação, exploração e cruzamento de diferentes formatos de dados (texto, imagem, áudio e vídeo); d) permitem visualizações avançadas dos dados; entre muitas outras (COSTA, 2016; LAGE; GODOY, 2008; SPANNAGEL; GLASER-ZUKUDA; SCHROEDER, 2005). Identificam-se estas características como possíveis vantagens da utilização deste tipo de *software*.

Elencar as desvantagens no uso destas ferramentas é algo que acompanhou o surgimento destes pacotes de *software*. Brent (1984) na década de 1980 já mencionava como desvantagem as

competências computacionais que os investigadores teriam de adquirir. Gibbs, Frieze, & Mangabeira (2002) afirmam que o *software* é menos útil para abordar questões de validade e confiabilidade nas ideias temáticas que emergem durante a análise de dados. Lage e Godoy (2008) elencam uma lista críticas que são colocadas ao uso de *software* na análise de dados qualitativos: a) possibilidade de perder o controlo no processo de codificação; b) confundir a *software* com a metodologia; c) encorajamento de estruturas de codificação complexas e detalhadas, resultando num excesso de codificação; d) aumento desnecessário da quantidade de dados recolhidos, levando ao risco do comprometimento da análise em profundidade; e) impossibilidade de comunicação entre sistemas (pacotes de *software*); f) levam os investigadores a utilizar um método particular de análise, de acordo com as características da ferramenta.

Poderíamos continuar a explicar as limitações e as potencialidades do uso destas ferramentas, contudo as características que as constituem atualmente conferem-lhes a credibilidade necessária para serem cada vez mais exploradas, tornando-as também mais robustas. Isto permite ao utilizador abdicar de realizar tarefas meramente “técnicas”, isto é, que não requerem um esforço intelectual e, por esse motivo, podem ser desempenhadas e, refira-se, largamente otimizadas pelo *software*. De forma análoga ao uso específico de ferramentas para a análise de dados, nenhum leitor acreditaria que este texto foi escrito com uso de uma máquina de datilografar.

## AGRADECIMENTOS

Este trabalho é apoiado financeiramente pelos Fundos Nacionais através da FCT - Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UID / CED / 00194/2013.

## REFERÊNCIAS

BAUGH, J.; HALLCOM, A. S.; HARRIS, M. E. Computer Assisted Qualitative Data Analysis Software: A Practical Perspective for Applied Research. **Revista Del Instituto Internacional de Costos**, n. enero/juni(6), 2010.

BAZELEY, P. **Qualitative data analysis with NVivo**. Thousand Oaks, California: SAGE Publication, 2007.

BRENT, E. Qualitative computing: Approaches and issues. **Qualitative Sociology**, v. 7, n. 1–

2, p. 34–60, 1984.

COSTA, A. P. Cloud Computing em Investigação Qualitativa: Investigação Colaborativa através do software webQDA. **Fronteiras: Journal of Social, Technological and Environmental Science**, v. 5, n. 2, p. 153–161, 12 dez. 2016.

COSTA, A. P.; AMADO, J. **Análise de Conteúdo suportada por software**. 1ª ed. Oliveira de Azeméis - Aveiro - PORTUGAL: Ludomedia, 2018.

COSTA, A. P.; LINHARES, R.; SOUZA, F. N. DE. Possibilidades de Análise Qualitativa no webQDA e colaboração entre pesquisadores em educação em comunicação. In: LINHARES, R.; FERREIRA, S. DE L.; BORGES, F. T. (Eds.). . **Infoinclusão e as possibilidades de ensinar e aprender**. Universidade Tiradentes, Aracaju - Brasil: Editora da Universidade Federal da Bahia, 2014. p. 205–215.

COSTA, A. P.; SOUZA, D. N. DE; SOUZA, F. N. DE. Trabalho Colaborativo na Investigação Qualitativa através das Tecnologias. In: SOUZA, D. N. DE; COSTA, A. P.; SOUZA, F. N. DE (Eds.). . **Investigação Qualitativa: Inovação, Dilemas e Desafios**. 1ª ed. Oliveira de Azeméis - Aveiro - PORTUGAL: Ludomedia, 2016. p. 105–127.

GIBBS, G. R.; FRIESE, S.; MANGABEIRA, W. C. The Use of New Technology in Qualitative Research. Introduction to Issue 3(2) of FQS. **Forum Qualitative Sozialforschung / Forum: Qualitative Social Research; Vol 3, No 2 (2002): Using Technology in the Qualitative Research Process**, v. 3, n. 2, 2002.

KAEFER, F.; ROPER, J.; SINHA, P. A Software-Assisted Qualitative Content Analysis of News Articles : Example and Reflections. **Forum Qualitative Sozialforschung**, v. 16, n. 2, 2015.

LAGE, M. C.; GODOY, A. S. O uso do computador na análise de dados qualitativos: questões emergentes. **RAM. Revista de Administração Mackenzie**, v. 9, n. 4, p. 75–98, 2008.

RIBEIRO, J. M.; BRANDÃO, C.; COSTA, A. P. Metodologia de Estudo de Caso em Saúde: Contributos para a sua Qualidade. In: OLIVEIRA, E. S.; BARROS, N.; RAIMUNDA SILVA



(Eds.). . **Investigação Qualitativa em Saúde: conhecimento e aplicabilidade**. 1<sup>a</sup> ed. Oliveira de Azeméis - Aveiro - PORTUGAL: Ludomedia, 2016. p. 143–160.

RICHARDS, L. Qualitative computing--a methods revolution? **International Journal of Social Research Methodology**, v. 5, n. 3, p. 263–276, jul. 2002.

SERÇE, F. C. et al. Online collaboration: Collaborative behavior patterns and factors affecting globally distributed team performance. **Computers in Human Behavior**, p. 1–14, 2010.

SOUZA, D. N. DE; COSTA, A. P.; SOUZA, F. N. DE. Desafio e inovação do estudo de caso com apoio das tecnologias. In: SOUZA, F. N. DE; SOUZA, D. N. DE; COSTA, A. P. (Eds.). . **Investigação Qualitativa: Inovação, Dilemas e Desafios (Volume 2)**. 1<sup>a</sup> ed. Oliveira de Azeméis - Aveiro - PORTUGAL: Ludomedia, 2015. p. 143–162.

SOUZA, F. N. DE; COSTA, A. P.; MOREIRA, A. **webQDA - Qualitative Data Analysis (versão 3.0)**AveiroMicro IO e Universidade de Aveiro, , 2016. Disponível em: <[www.webqda.net](http://www.webqda.net)>

SPANNAGEL, C.; GLASER-ZUKUDA, M.; SCHROEDER, U. Application of Qualitative Content Analysis in User-Program Interaction Research. **Forum Qualitative Sozialforschung/Forum: Qualitative Social Research**, v. 6, n. 2, 2005.